Pain Medicine
Pain Medicine 2012; 13: 1347–1357
Wiley Periodicals, Inc.

Use of Machine-Learning Classifiers to Predict Requests for Preoperative Acute Pain Service Consultation

Patrick J. Tighe, MD,* Stephen D. Lucas, MD,* David A. Edwards, MD, PhD,* André P. Boezaart, MD, PhD,*† Haldun Aytug, PhD,‡ and Azra Bihorac, MD*

Departments of *Anesthesiology and

†Orthopaedic Surgery, University of Florida College of Medicine, Gainesville, Florida;

‡Department of Information Systems and Operations Management, University of Florida Warrington College of Business, Gainesville, Florida, USA

Reprint requests to: Patrick J. Tighe, MD, Department of Anesthesiology, University of Florida College of Medicine, PO Box 100254, 1600 SW Archer Road, Gainesville, FL 32610-0254, USA. Tel: 352-273-6575; Fax: 352-392-7029; E-mail: ptighe@anest.ufl.edu.

Potential conflicts of interest: Dr. Boezaart receives financial compensation from Teleflex Medical for consultancy and lectures, as well as royalty payments for the invention of the stimulating catheter. Other authors named do not have a possible conflict of interest to declare.

Funding source: Supported in part by Grant UL1 RR029890 through NIH Clinical and Translational Science Awards Program, University of Florida Clinical and Translational Science Institute, Gainesville, Florida. Dr. Bihorac was supported by the National Institutes of Health, NIGMS award K23GM087709. Dr. Tighe was supported by the National Institutes of Health, NIGMS award K23 GM102697.

Presentation/publication at meetings: This material was presented in abstract form at the American Society of Anesthesiologists Annual Meeting, October 15–19, 2011.

Abstract

Objective. The purpose of this project was to determine whether machine-learning classifiers could predict which patients would require a preoperative acute pain service (APS) consultation.

Design. Retrospective cohort.

Setting. University teaching hospital.

Subjects. The records of 9,860 surgical patients posted between January 1 and June 30, 2010 were reviewed.

Outcome Measures. Request for APS consultation. A cohort of machine-learning classifiers was compared according to its ability or inability to classify surgical cases as requiring a request for a preoperative APS consultation. Classifiers were then optimized utilizing ensemble techniques. Computational efficiency was measured with the central processing unit processing times required for model training. Classifiers were tested using the full feature set, as well as the reduced feature set that was optimized using a merit-based dimensional reduction strategy.

Results. Machine-learning classifiers correctly predicted preoperative requests for APS consultations in 92.3% (95% confidence intervals [CI], 91.8–92.8) of all surgical cases. Bayesian methods yielded the highest area under the receiver operating curve (0.87, 95% CI 0.84–0.89) and lowest training times (0.0018 seconds, 95% CI, 0.0017–0.0019 for the NaiveBayesUpdateable algorithm). An ensemble of high-performing machine-learning classifiers did not yield a higher area under the receiver operating curve than its component classifiers. Dimensional reduction decreased the computational requirements for multiple classifiers, but did not adversely affect classification performance.

Conclusions. Using historical data, machine-learning classifiers can predict which surgical cases should prompt a preoperative request for an APS consultation. Dimensional reduction improved computational efficiency and preserved predictive performance.

Key Words. Machine Learning; Pain; Surgery; Anesthesia; Acute Pain Service

Introduction

The modern acute pain service (APS) requires a specialized organizational structure involving additional resources in the form of nurses, technicians, and clerks specifically educated to meet the needs of the APS [1]. Additionally, to optimize resident education and patient flow, many academic medical centers provide APS services in a dedicated space, often referred to as a "block room" [2]. The optimal use of these relatively costly resources depends on the appropriate and timely referral by operating surgeons. However, due to the lack of evidence-based guidelines for the perioperative need for APS consultations, the referral pattern is often subjected to the randomness of decisions by individual referring physicians, and thus may appear to be highly unpredictable [3]. This apparent inability to forecast the request for APS services in a timely fashion may prevent providers from using cost-effective strategies, such as parallel processing, and may minimize efficiency improvements obtained with a regional anesthesia block room model [2,4,5].

On the other hand, the experience of industries, such as insurance or risk-management companies, that depend on the effective prediction of future outcomes highlights the increasing power of predictive analytics as an emerging forecasting tool. Predictive analytics encompass a variety of statistical techniques to analyze highly dimensional data sets with large numbers of variables, identify patterns based upon given sets of outcomes, and use these patterns to successfully forecast future events with iterative improvements [6]. Among others, machine learning, a branch of artificial intelligence, is gaining popularity, as it not only uses advanced statistical methods for regression and classification, but also emulates human cognition and the ability to learn from training examples to predict future events [7]. Since the inception of the Health Information Technology for Economic and Clinical Health Act initiatives, more hospitals and clinics are integrating electronic medical records (EMRs) into their health information processing and storage, thereby enabling the application of predictive analytics for the forecasting of events in the health care industry [8]. Consequently, the Agency for Healthcare Research and Quality has promoted the use of health information technology, including clinical decision support systems, to advance the quality of patient care [9].

In this retrospective cohort study, we tested the hypothesis that, by using readily available information in EMRs, predictive analytics could reliably and cost-efficiently predict the request for an APS consultation ahead of the actual request. In addition, we compared the performance and efficiency of different types of machine learning classifiers (MLC) when applied to this classification task. Finally, we examined the role of ensemble classification techniques and dimensional reduction in machine learning classification to forecast APS consultation requests.

Methods

Study Population

The University of Florida Institutional Review Board approved this study. All in-hospital surgical cases performed between January 1, 2010 and June 30, 2010 at the University of Florida and Shands Hospital (UF & Shands), a tertiary academic center, were reviewed if they were posted on the operating room list and, hence, were included in the EMRs database (Centricity Perioperative Management [CPM], GE Healthcare, Waukesha, WI, USA). A preoperative request for an APS consultation, in practice, leads to the patient being sent to a preoperative block room for preoperative assessment and evaluation for placement of a nerve block.

Outcomes and Input Attributes

Only those attributes available preoperatively in the operating room schedule and CPM database were included. This limitation mimicked the preoperative decision point that would be available to operating room schedulers and resource assignment administrators. Input data included 11 attributes: operating room location, patient status description (inpatient, outpatient, or postoperative admission), surgeon, anesthesiologist, primary and secondary Current Procedural Terminology (CPT) codes, day of week, scheduled surgery starting time, postoperative care location, patient age, and estimated duration of surgery.

For input data, continuous variables were nominalized to ensure broad application to differing classifiers (see later). Primary operating surgeon and primary attending anesthesiologist identities and primary and secondary CPT codes were included as nominal values. For purposes of demonstrating the distribution of outcome predictions across procedural groups, CPT codes were also grouped into broader anatomic surgical categories; such groupings were not used for classifier development or testing, but, rather, for demonstration of outcome predictions across surgical categories (see Appendix 1). The day of the week for each procedure was included to account for cases scheduled outside of a surgeon's standard OR time block, but which may still be considered elective. Likewise, the location of the operating room was included as a categorical variable to account for procedures performed outside of the surgeon's or customary venue of the procedure. Scheduled starting time was nominalized into five categories: 0700–1159, noon–1500, 1501–1800, 1801–2100, and 2101–0659, as these time intervals represented differing practice patterns and case volumes for the APS. Postoperative destinations included the post-anesthesia care unit (PACU), surgical intensive care unit (ICU), pediatric ICU, neonatal ICU, medical ICU, cardiac ICU, burn ICU, and neurosurgical ICU. Patient age was categorized (according to empirically derived groups) as 1 year or less, 1–2 years, 3–5 years, 6–10 years, 11–15 years, 16–20 years, and 21–30 years, with 10-year increments up to 100. The estimated surgery times were defined as 1–30 minutes, 31–60 minutes, 61–90 minutes, 91–120 minutes, 121–180 minutes, 181–240 minutes, 241–300 minutes, 301–360 minutes, 361–420 minutes, 421–480 minutes, and >8 hours.

The primary outcome was the preoperative APS consultation request; this was considered a binary outcome.

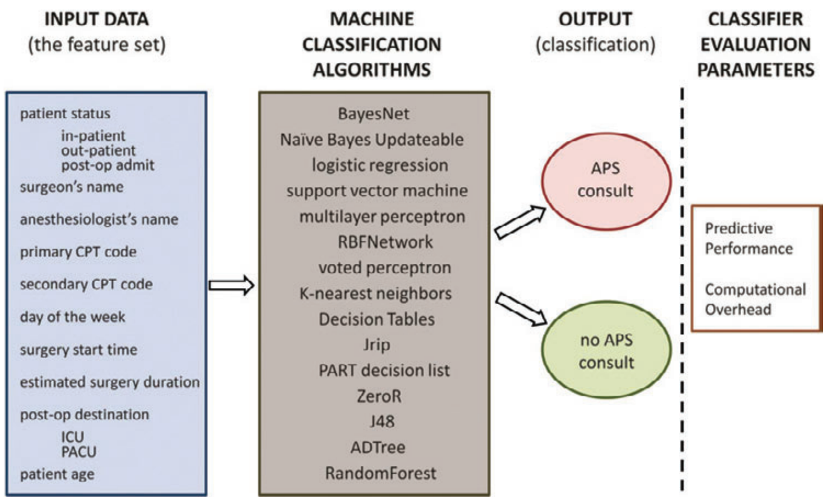


Figure 1 Overview of machine learning classifier (MLC) design, implementation, and testing. A series of inputs, called the feature set, are entered into a series of MLCs. The classifiers then render an output prediction, based upon prior training, using a feature set with known outcomes. In this study, classifiers were evaluated using metrics of both predictive performance as well as computational requirements. CPT = Current Procedural Terminology; ICU = intensive care unit; PACU = post-anesthesia care unit; APS = acute pain service.

MLC Selection, Performance, and Optimization

The main goal of machine learning research is to create machines (computers) that can learn [6]. Classification, one of the most common machine learning tasks, has attracted much attention. Classification algorithms induce a classifier by using training examples (i.e., data that are prelabeled by a supervisor). There are many ways to represent a classifier, with common representations that include "if-then rules," "decision trees," "networks" (such as neural networks), "mathematical functions" (such as linear functions), and many others. Regardless of the method of classifier training used, the goal of a classifier remains the same: determining to which group a new observation should be assigned by using the information contained in prior observations for guidance. In this project, we used supervised machine learning techniques to develop classifiers, which we refer to as machine-learning classifiers (MLC), for predicting whether a given surgical case will require an APS consultation.

The overall mechanism for MLC creation is a repetitive process divided into two phases. In the first phase, or training phase, a set of data with known outcomes is used to train the classifier. In the second phase, or testing phase, the performance of the classifier is evaluated by using it to make predictions on previously unseen data. Because the classifier is applying lessons learned from the training data set to a new, previously unseen set of instances in need of classification, MLC performance may decrease to reflect a more realistic scenario. We combined the testing and training phases using stratified 10-fold cross-validation by segmenting the entire data set into 10 "folds" (Figure 1).

We used different MLC types to represent a variety of classification methodologies (Table 1) [6,10]. All classifiers, with the exception of three, were able to complete the training and testing phases. The support vector machine, multilayer perceptron, and logistic regression schemes each required greater than 10 minutes of dedicated central processing unit (CPU) time to partially train the MLC, and, thus, were excluded from further study. The following parameters were used to compare the performance of MLCs: area under the receiver operating curve (AUC), percent correct, sensitivity, specificity, and computational requirements. The AUC was selected as the primary performance metric due to its comprehensive description of overall classifier performance [11,12]. "Percent correct," or accuracy, represents a traditional measurement of classifier performance, but falsely assumes uniform class distribution across the study sample [11]. Computational requirements were determined using the CPU time required for model training. All models were tested using Weka 3.6.2 (University of Waikato, New Zealand).

In order to visually demonstrate a decision tree, we applied a separate dedicated decision tree algorithm (JMP Pro 9, SAS Institute, Cary, NC, USA). As a decision tree incorporating all variables would be prohibitively challenging to visualize given the large number of levels inherent to our data set, this example used a separate sample of variables. All attributes were restricted to variables with

Table 1 Description of machine learning classification schemes

Type of Classifier	Examples	Description
Bayesian	BayesNet Naïve Bayes (updateable)	Apply Bayesian principles of conditional probability to classification. Naive Bayesian approaches presume that input factors are independent of one another.
Function-based	Logistic regression Support vector machine Multilayer perceptron Radial basis frequency network Voted perceptron	Creates a separating surface using linear, logistic, or kernel-based methods.
Lazy	K-nearest neighbor	Maps outcomes onto a plot, and classifies outcomes based upon proximity to neighboring outcomes with known correct classification via majority rules.
Rule-based	Decision tables Propositional rule learner (JRip) PART decision list ZeroR	Create rules to assign outcomes into the correct classification grouping, either simply or via iteration.
Decision trees	J48 AD Tree Random Forest	Create hierarchy of rules for classification, using an "If this . . . then that . . ." paradigm.

less than 30 levels, taking into account both algorithm constraints and visualization limitations; CPT anatomic groupings, patient status, day of week, scheduled start time, postoperative location, patient age group, and estimated surgical duration time were used for this simplified visualization (Figure 2).

Predictive Model Optimization

To further optimize predictive model performance and decrease computational requirements, we used two strategies: ensemble development and dimensional reduction. We combined MLC types with the highest AUC from each methodology group into an ensemble meta-classifier using the Vote schema, which averages the probability estimates of component base classifiers to render a single classification per instance [13]. In the second optimization step, we performed dimensional reduction to identify key attributes in the prediction model [14,15]. We employed a correlation-based feature subset selection (CFS) using the greedy-stepwise search pattern with 10-fold cross-validation [16]. This dimensional reduction strategy identifies those attributes that are highly correlated with the separating class, yet poorly correlated with the other attributes. The resulting "merit" of an attribute results from the following equation:

$$M_S = k^* r_{cf} \Big/ \sqrt{(k + k\,(k-1)\,r_{ff})}\,,$$

where M is the merit of a subset of feature set S, which contains k number features; r_{cf} is the average correlation between a feature and a class; and r_{ff} is the average intercorrelation between features. Thus, the numerator indicates how predictive a feature set is, and the denominator indicates the redundancy among features.

The greedy-stepwise search pattern identifies a baseline set of attributes, and then iteratively adds or deletes attributes to improve the correlation score within the CFS selection schema [10,15]. The rank and merit of each attribute was reported based on the results of the CFS analysis. Only those attributes contributing to the majority of cross-validated folds were used to further test the MLC performance in a dimensionally reduced data set.

Statistical Analysis

MLC performance measures were compared using non-parametric analysis of variance. Post hoc comparisons were conducted using the Steel or Dunn's method, as appropriate. Pairwise comparisons of before and after dimensional reduction were performed using Wilcoxon's method. Statistical significance was set, with alpha <0.05. Power analysis, using a one-sample mean per classifier and testing a difference to detect an AUC of 0.1 with a standard deviation of 0.1 and sample size of 10 runs, suggested a power of 0.8. Statistical analysis of classifier performance comparisons was conducted using JMP 9 (SAS Institute).

Results

Model Description

A total of 9,860 surgical cases were included in the analysis. Sixteen percent of them had preoperative APS consultations, and were evaluated in the block room immediately prior to surgery. Over the included 6-month time frame, this represents an average of 13 consultations per week day, or nine consultations per day if averaged across all days. There were 43 separate operating locations, ranging from 0.04% to 5.1% of cases. Forty-two percent of patients were inpatient, 21% outpatient, and 37% represented patients admitted to the hospital following surgery. A total of 186 surgeons were included, with none accounting for more than 3% of the case volume. Likewise, 77 attending anesthesiologists were included, ranging from 0.01% to 6% of all cases. A total of 1,514 primary CPT codes were used in conjunction with 833 secondary CPT codes. Cases were performed on every day of the week, 94% of which were performed Monday–Friday. Surgery started between 0700 hours and noon in 66% of cases, dropping to 2% between 1801 and 2100 hours, and 3% between 2100 and 0700 hours. The most common postoperative destination was the PACU for 73% of patients, with 0.2–6% of patients transferred to each type of ICU. Ages ranged from less than 1 year to 100 years. Most procedures lasted less than 2 hours and over 25% less than 1 hour.

Performance of Predictive Models

Using all 11 attributes in the full model, the MLCs correctly predicted preoperative requests for APS consultations in 92.3% (95% confidence intervals [CI], 91.8–92.8) of all surgical cases. There was a statistically significant difference between the different types of MLCs in the percentage of correct predictions ($P < 0.0001$), ranging from 84.2% (95% CI, 84.2–84.2), with ZeroR, to 94.3% (95% CI, 93.7–94.9) with the Voted Perceptron. The AUC ranged from 0.5, with ZeroR, to 0.95 (95% CI, 0.95–0.96) for both BayesNet and NaiveBayesUpdateable algorithms, with a statistically significant difference in the AUC among the classifiers ($P < 0.0001$). A representation of classifier predictions rendered by BayesNet across different anatomic surgical groupings of the primary CPT code is given in Figure 3.

The BayesNet and NaiveBayesUpdateable algorithms returned the highest classifier sensitivities (0.87, 95% CI 0.84–0.89 and 0.87, 95% CI 0.84–0.89, respectively). The ZeroR classifier had the lowest sensitivity at zero. Contrariwise, ZeroR was the most specific classifier, with a specificity of 1; and BayesNet (0.94, 95% CI 0.93–0.95), NaiveBayesUpdateable (0.94, 95% CI 0.93–0.95), and IBk (0.95, 95% CI 0.94–0.95) were the least specific. Overall, there were significant differences between classifiers in both sensitivity ($P < 0.0001$) and specificity ($P < 0.0001$).

The Voted Perceptron required the highest CPU time for training, at 38.7 (95% CI, 38.2–39.2) seconds, and NaiveBayesUpdateable required the least, at 0.0018 (95% CI, 0.0017–0.0019) seconds. Overall, there was a significant difference among classifiers in the time required for CPU training ($P < 0.0001$) (Table 2).

Performance of Optimized Predictive Models

The MLCs with the highest AUC from each group were included in the EnsembleVote meta-classifier. The EnsembleVote demonstrated an AUC of 0.95 (95% CI, 0.95–0.96), percent correct of 94.7% (95% CI, 94.2–95.3), sensitivity of 0.82 (95% CI, 0.79–0.85), specificity of 0.97 (95% CI, 0.97–0.97), and CPU training time of 3.69 seconds (95% CI, 3.60–3.77). No significant difference was demonstrated between the EnsembleVote AUC and the AUC of the individual components of the Ensemble. The EnsembleVote had significantly higher CPU training time requirements when compared with almost all classifiers (Table 3).

The CFS dimensional reduction algorithm recognized six separate attributes that highly correlated with the outcome classification: attending anesthesiologist, surgeon, primary and secondary CPT code, location of operating room, and scheduled start time. Each of these attributes contributed to classification in all cross-validated folds (Table 4). Here, the folds denote each of the 10 cross-validated folds used in the dimensional reduction algorithm, demonstrating that the six relevant attributes were found to be of significance in all tested folds, while the attributes that were ranked 7 to 11 had low merit scores, suggesting a poor independent correlation with outcome and/or high levels of correlation with other attributes. Although dimensional reduction did not lead to significant changes in MLC performance, it did significantly reduce CPU training time for all models, ranging from an 11% to 38% reduction (Figure 4).

Discussion

Our study demonstrated that predictive analytics methodology MLCs and reasonable computational resources can be successfully used to predict a request for a preoperative APS consultation in 92% of surgical cases based upon readily available data in EMRs. While this study was restricted entirely to retrospective data for training and testing of the MLCs, the cross-validated results suggest that MLCs may be able to prospectively predict those patients for whom an APS consult is likely to be requested. To the best of our knowledge, this is the first published study that has used a nontraditional predictive analytic methodology to support clinical decisions regarding pain service-related clinical resource allocation. This success becomes more noteworthy when considering the failure of traditional statistical methods to render predictions by utilizing realistic computational resources, as we discovered when we attempted to employ logistic regression in our study, but failed due to high processing times. We found that most classifiers achieved a high AUC, percent correct, and specificity scores regardless of the underlying classification methodology, with the major difference in computational requirements for performance.

An increasing number of investigations using quantitative sensory testing, genetic biomarkers, and psychometric evaluations to assist with predictions of postoperative pain presage an era in which the clinician will need to transform larger quantities of increasingly complex information into real-time decisions [17–20]. In recognition of this, the Centers for Medicare and Medicaid Services has required

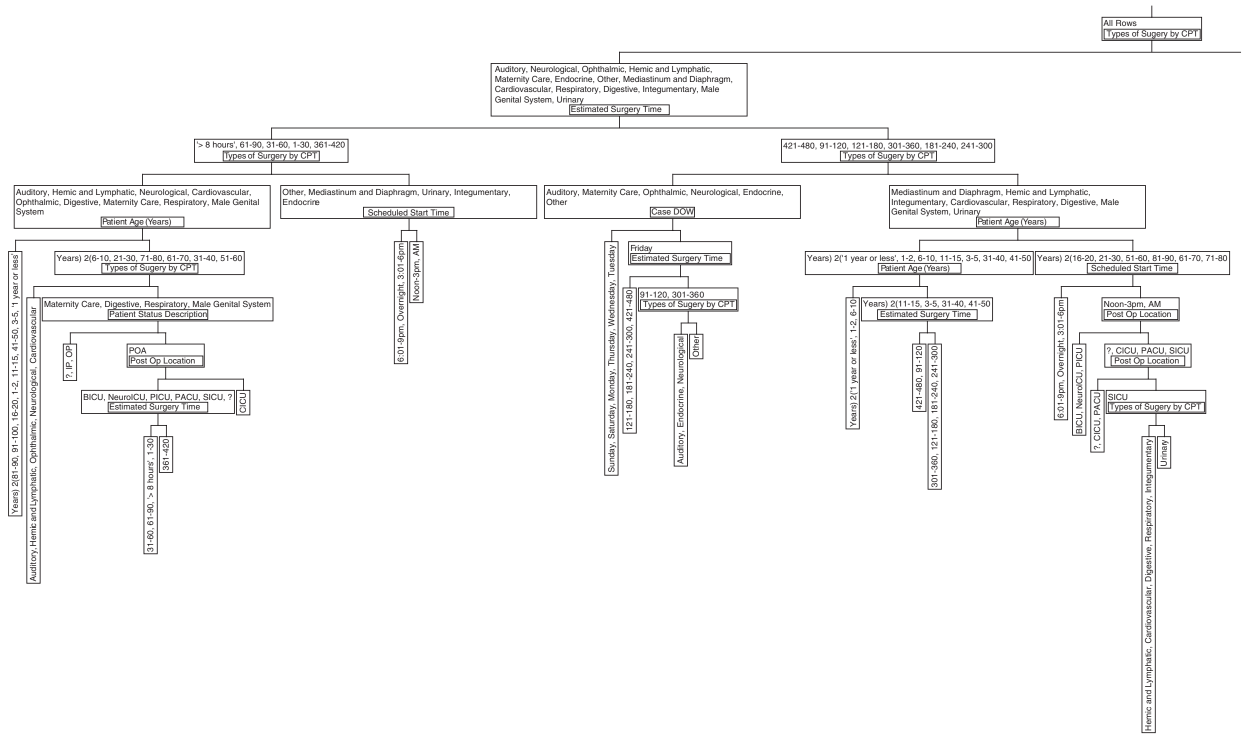
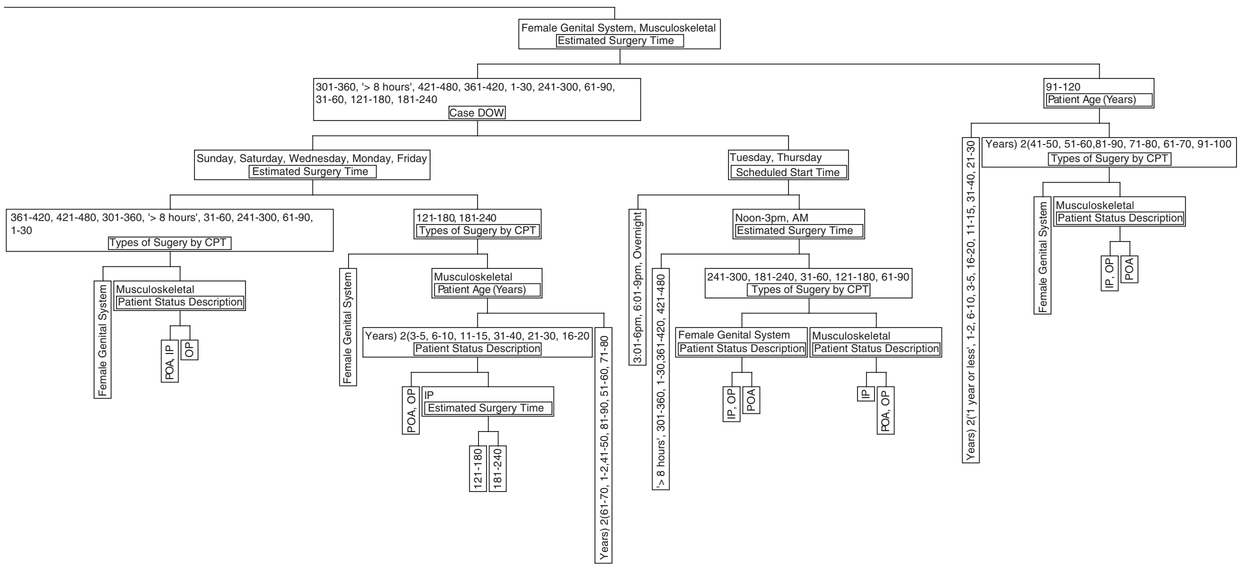


Figure 2 Example of a decision tree for forecasting APS consult requests. This decision tree used only a subset of variables in order to avoid difficulty with visualizing the branch and leaf pattern inherent to variables with hundreds of levels. Attributes were restricted to variables with less than 30 levels; and included patient status, and CPT groupings by anatomic site, day of week, scheduled start time, postoperative location, patient age group, and estimated surgical duration. The 10-fold cross-validated ROC was 0.89.

that EMRs must demonstrate decision support service capabilities to qualify for EMR-based performance incentives [8,21]. Our results suggest that predictive analytics techniques, such as MLCs, may offer considerable assistance as clinical decision support tools, but must be compared and validated using multiple metrics of classifier performance. The results of this study complement previous work demonstrating the potential utility of MLC to forecast the need for a postoperative rescue femoral nerve block following anterior cruciate ligament repair [22].

Our particular APS involves the use of a preoperative "block room" consisting of a dedicated space in the preoperative holding area, and specialized preoperative nurses who assist with patient preparation, sedation, and block placement. Although the block room offers trainees a more protected and less time-pressured environment to place nerve blocks, their ability to decrease total perioperative costs is debatable [2,5,23,24]. The automated prediction, of which patients are expected to have an APS consultation and be sent to the block room ahead of the actual consultation request by the surgeon, can help with more appropriate advance staffing, potentially decreasing the cost of a block room.

Interestingly, the high AUC scores across multiple MLCs point to a highly patterned practice within our health care system. This might suggest that practitioners should be able to predict which patients need a preoperative APS consultation with a fair degree of certainty. At our own institution, our APS team spends as much as 45 minutes each day reviewing over 150 cases across three facilities to determine which patients are likely to require preoperative APS consultations. Despite this investment of time, our system still errs, with up to 1–2 patients per day affected. Examples leading to such errors include surgeon requests for a nerve block at the last minute or a late addition of a patient to the operating room schedule that includes a delayed discussion with the APS team concerning nerve block placement. Such errors can be magnified when the involved physicians are trainees or new to local practice patterns. Our results suggest that MLCs may provide clinical decision support using information already available, potentially decreasing both time requirements and decision-making errors. Future research will need to compare MLCs and human classification performance in this problem set, as well as determine whether human predictions can amplify MLC performance.



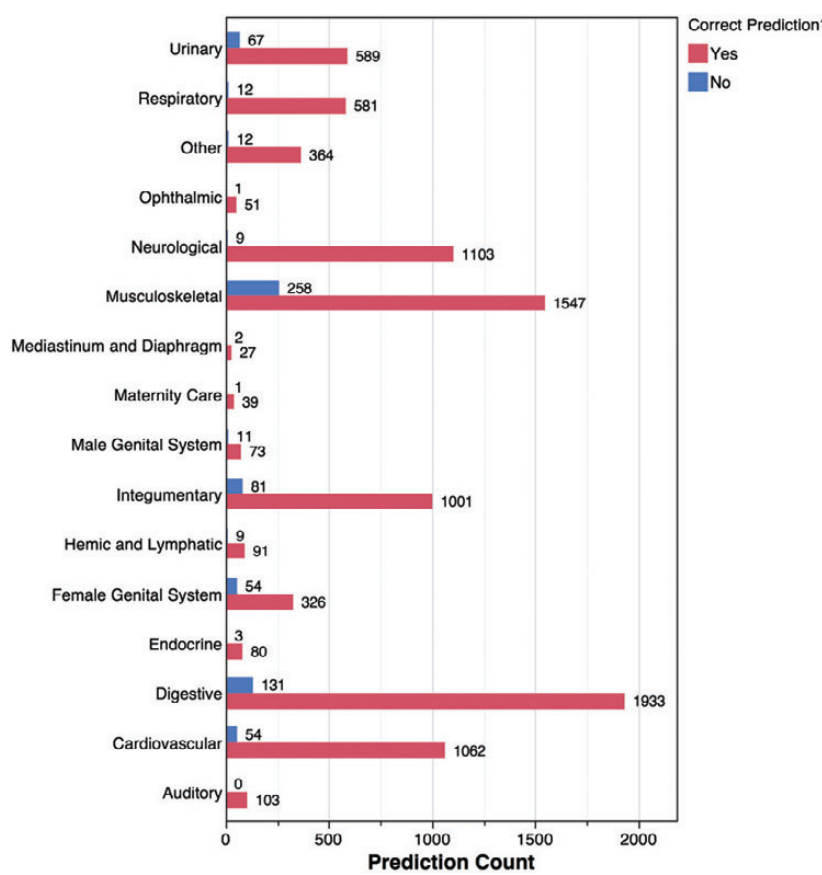


Figure 3 Distributions of correct and incorrect predictions by anatomic groupings of CPT codes. Both correct and incorrect predictions are listed for each CPT code.

Table 2 Summary of machine learning classifier performance

Classifier	Percent Correct	Area under ROC	Sensitivity	Specificity	CPU Training Time (second)
BayesNet	92.8	0.95	0.87	0.94	0.01
NaiveBayesUpdateable	93.0	0.95	0.86	0.94	0.0018
RBFNetwork	93.9	0.95	0.81	0.96	0.27
VotedPerceptron	94.3	0.87	0.75	0.98	38.7
IBk	92.0	0.91	0.78	0.95	0.0005
DecisionTable	91.6	0.88	0.68	0.96	2.3
JRip	93.4	0.88	0.78	0.96	3.47
PART	91.4	0.88	0.54	0.98	3.22
ZeroR	84.2	0.50	0.00	1.00	0.0005
RandomForest	93.5	0.95	0.71	0.98	1.09
ADTree	92.5	0.89	0.69	0.97	0.073
J48	92.5	0.88	0.68	0.97	0.20
EnsembleVote	94.7	0.95	0.82	0.97	3.69

Table 3 Comparison of classifier performance between EnsembleVote and naive classifiers*

Native Classifier	Native Classifier aROC	EnsembleVote aROC	*P* Value
BayesNet	0.95	0.95	1
NaiveBayesUpdateable	0.95	0.95	1
RBFNetwork	0.95	0.95	1
VotedPerceptron	0.87	0.95	<0.0001
IBk	0.91	0.95	0.2
Decision table	0.88	0.95	0.002
JRip	0.88	0.95	0.0005
PART	0.88	0.95	0.002
ZeroR	0.50	0.95	<0.0001
RandomForest	0.95	0.95	1
ADTree	0.89	0.95	0.003
J48	0.88	0.95	0.001

Native Classifier	Native Classifier CPU Training Time (second)	EnsembleVote CPU Training Time (second)	*P* Value
BayesNet	0.01	3.69	0.0002
NaiveBayesUpdateable	0.0018	3.69	<0.0001
RBF Network	0.27	3.69	0.1
VotedPerceptron	38.7	3.69	1
IBk	0.0005	3.69	<0.0001
DecisionTable	2.3	3.69	1
JRip	3.47	3.69	1
PART	3.22	3.69	1
ZeroR	0.0005	3.69	<0.0001
RandomForest	1.09	3.69	0.6
ADTree	0.073	3.69	0.002
J48	0.20	3.69	0.02

* Resulting *P* values are listed following nonparametric comparisons using the Dunn's method between the Naive and Ensemble Classifiers.

Table 4 Results of the correlation feature subset selection method of dimensional reduction

Rank	Attribute	Average Merit	Number of Folds
1	Anesthesiologist	0.087 ± 0.001	10/10
2	Surgeon	0.106 ± 0.001	10/10
3	Primary CPT code	0.108 ± 0.001	10/10
4.2 ± 0.4	OR number	0.109 ± 0.001	10/10
5.2 ± 0.75	Secondary CPT code	0.109 ± 0.001	10/10
5.6 ± 0.49	Scheduled start time	0.11 ± 0.001	10/10
7	Patient status description	0.109 ± 0.001	0/10
8.4 ± 0.49	Estimated duration of surgery	0.105 ± 0.001	0/10
8.6 ± 0.49	Postoperative destination	0.105 ± 0.001	0/10
10	Patient age	0.101 ± 0.001	0/10
11	Case day of week	0.099 ± 0.001	0/10

Rather than seeking improvement through added complexity, optimization techniques employing dimensional reduction may offer substantial improvements in computational efficiency and preserve classification ability. No classifiers in our study suffered a statistically significant decrease in the AUC when moving from a full to reduced set of input data. However, multiple classifiers, including those with a high AUC, such as BayesNet, RBFNetwork, and EnsembleVote, exhibited statistically significant decreases in computational requirements. Although the proportional decrease in time requirements was as high as 38%, the magnitude of differences was, nevertheless, small. Such differences would likely be clinically insignificant unless implemented in environments focused on efficient CPU utilization, such as with mobile devices. On the other hand, by combining several unique classifiers into a single meta-classifier, EnsembleVote, we achieved only incremental improvements in the AUC that were not statistically significant.

Although our findings are derived at the single institution level, hence reflecting an institutional practice pattern where continuous and single-injection peripheral, neuraxial, and paraneuraxial blocks are highly accepted

Figure 4 Comparison of area under the receiver operating curve (ROC) and CPU training time requirements for each classifier using full and reduced feature sets. For the area under the ROC, statistically significant differences between full and reduced feature sets were found only for the IBk classifier. Statistically significant differences in computational training times between the full and reduced feature sets were found for BayesNet, RBFNetwork, DecisionTable, ADTree, and EnsembleVote. CPU = central processing unit; ROC = receiver-operating curve. *Starred items denote differences between full and reduced feature set classifier results significant at the 0.05 level.

and supported by surgeons, anesthesiologists, patients, and administrators, the methodology is applicable globally. Notably, the CFS method of dimensional reduction does not offer coefficients, likelihood estimates, or other traditional metrics denoting the relative importance of a feature. Rather, it achieves the same goal of a stepwise approach to regression by calculating the correlation of a given feature as a function of other attributes as well as its class dependency. This approach was highly effective in our study in both reducing the number of features to improve processing efficiency and maintaining accurate classifications. Those attributes with high merit scores may warrant closer examination in external applications of our methodology.

One of the study limitations is the lack of patient factors in the data source used for MLC training. Ideally, MLCs would be able to predict which patients would benefit from a preoperative APS consultation, rather than limiting the question to who would be referred for consultation. To optimally predict which patients would benefit from an APS consultation, MLCs would need to account for the presence and timing of anticoagulation, physiologic parameters (i.e., severe emphysema, aortic stenosis), patient questions, and patient preferences after discussion of the risks, benefits, and alternatives. These requirements, of course, are in addition to the patient and procedural-specific outcome data for regional anesthetics. Ultimately, the widespread incorporation of information from EMRs will allow for the application of predictive analytics for more complex and automated clinical decision processes.

Conclusions

In conclusion, although the complex nature of pain renders special challenges to any predictive efforts, the ability of predictive analytics techniques to efficiently manage highly dimensional data offers novel opportunities to provide clinical predictors and circumvent the limitations inherent in traditional, logistic regression-based factor analysis. Future work is necessary to refine the many methodological issues, to identify the relevant data inputs, and to translate both data and its analysis into improved postoperative pain control.

Acknowledgments

The authors wish to thank Esperanza Olivo for her editorial support on this paper.

References

1 Rawal N. Organization, function, and implementation of acute pain service. Anesthesiol Clin North America 2005;23:211–25.

2 Drolet P, Girard M. Regional anesthesia, block room and efficiency: Putting things in perspective. Can J Anaesth 2004;51:1–5.

3 Masursky D, Dexter F, Mccartney CJ, Isaacson SA, Nussmeier NA. Predicting orthopedic surgeons' preferences for peripheral nerve blocks for their patients. Anesth Analg 2008;106:561–7.

4 Friedman DM, Sokal SM, Chang Y, Berger DL. Increasing operating room efficiency through parallel processing. Ann Surg 2006;243:10–4.

5 Armstrong KPJ, Cherry RA. Brachial plexus anesthesia compared to general anesthesia when a block room is available. Can J Anaesth 2004;51:41–4.

6 Witten I, Frank E. Data Mining: Practical Machine Learning Tools and Techniques, 2nd edition. San Francisco, CA: Morgan Kaufmann; 2005.

7 Bose I, Mahapatra R. Business data mining-a machine learning perspective. Inf Manage 2001;39:211–25.

8 Centers for Medicare & Medicaid Services. CMS medicare and medicaid EHR incentive programs. Milestone timeline. Available at: https://www.cms.gov/EHRIncentivePrograms/Downloads/EHRIncentProgtimeline508V1.pdf (accessed January 18, 2012).

9 Lyman JA, Cohn WF, Bloomrosen M, Detmer DE. Clinical decision support: Progress and opportunities. J Am Med Inform Assoc 2010;17:487–92.

10 Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning. New York: Springer Verlag; 2008.

11 Ben-David A. About the relationship between ROC curves and Cohen's kappa. Eng Appl Artif Intell 2008;21:874–82.

12 Metz CE. ROC analysis in medical imaging: A tutorial review of the literature. Radiol Phys Technol 2008;1:2–12.

13 Kittler J, Hatef M, Duin R, Matas J. On combining classifiers. IEEE Trans Pattern Anal Mach Intell 1998;20:226–39.

14 van der Maaten LJP, Postma EO, van den Herik HJ. Dimensionality reduction: A comparative review (Technical report, TiCC-TR 2009-005). The Netherlands: Tillburg University; 2009.

15 Jimenez LO, Landgrebe DA. Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data. IEEE Trans Syst Man Cybern 1998;28:39–54.

16 Hall MA. Correlation-based feature selection for machine learning. Dissertation. The University of Waikato, Hamilton, New Zealand; April 1999. Available at: http://www.cs.waikato.ac.nz/~mhall/thesis.pdf (accessed January 18, 2012).

17 Yarnitsky D, Crispel Y, Eisenberg E, et al. Prediction of chronic post-operative pain: Pre-operative DNIC testing identifies patients at risk. Pain 2008;138:22–8.

18 Janssen KJ, Kalkman CJ, Grobbee DE, et al. The risk of severe postoperative pain: Modification and validation of a clinical prediction rule. Anesth Analg 2008;107:1330–9.

19 Werner MU, Mjöbo HN, Nielsen PR, Rudin A. Prediction of postoperative pain: A systematic review of predictive experimental pain studies. Anesthesiology 2010;112:1494–502.

20 Kalkman CJ. Preoperative prediction of severe postoperative pain. Pain 2003;105:415–23.

21 Centers for Medicare & Medicaid Services. CMS finalizes definition of meaningful use of certified electronic health records (EHR) technology. Released July 16, 2010. Available at: https://www.cms.gov/apps/media/press/factsheet.asp?Counter=3794&intNumPerPage=10&checkDate=&checkKey=&srchType=1&numDays=3500&srchOpt=0&srchData=&keywordType=All&chkNewsType=6&intPage=&showAll=&pYear=&year=&desc=&cboOrder=date (accessed April 10, 2012).

22 Tighe P, Laduzenski S, Edwards D, et al. Use of machine learning theory to predict the need for femoral nerve block following ACL repair. Pain Med 2011;12:1566–75.

23 Mariano ER, Chu LF, Peinado CR, Mazzei WJ. Anesthesia-controlled time and turnover time for ambulatory upper extremity surgery performed with regional vs general anesthesia. J Clin Anesth 2009;21:253–7.

24 Dexter F, Epstein RH. Operating room efficiency and scheduling. Curr Opin Anaesthesiol 2005;18:195–8.

Appendix 1

Type of Surgery	CPT Range
General	10021–10022
Integumentary	10040–19499
Musculoskeletal	20000–29999
Respiratory	30000–32999
Cardiovascular	33010–37799
Hemic and lymphatic	38100–38999
Mediastinum and diaphragm	39000–39599
Digestive	40490–49999
Urinary	50010–53899
Male genital	54000–55899
Reproductive	55920–55980
Female genital	56405–58999
Maternity care	59000–59899
Endocrine	60000–60999
Neurological	61000–64999
Ophthalmic	65000–68999
Auditory	69000–69999